Optional ERIC Coversheet — Only for Use with U.S. Department of Education Grantee Submissions

This coversheet should be completed by grantees and added to the PDF of your submission if the information required in this form **is not included on the PDF to be submitted**.

INSTRUCTIONS

- Before beginning submission process, download this PDF coversheet if you will need to provide information not on the PDF.
- Fill in all fields—information in this form **must match** the information on the submitted PDF and add missing information.
- Attach completed coversheet to the PDF you will upload to ERIC [use Adobe Acrobat or other program to combine PDF files]—do not upload the coversheet as a separate document.
- Begin completing submission form at https://eric.ed.gov/submit/ and upload the full-text PDF with attached coversheet when indicated. Your full-text PDF will display in ERIC after the 12-month embargo period.

GRANTEE SUBMISSION REQUIRED FIELDS

Title of article, paper, or other content

Rethinking Universal Suspension for Severe Student Behavior

All author name(s) and affiliations on PDF. If more than 6 names, ERIC will complete the list from the submitted PDF.

Last Name, First Name	Academic/Organizational Affiliation	ORCID ID
Hinze-Pifer, Rebecca	Stanford University	
Sartain, Lauren	University of Chicago	

Publication/Completion Date—(if *In Press,* enter year accepted or completed) 2018

Check type of content being submitted and complete one of the following in the box below:

- ☑ If article: Name of journal, volume, and issue number if available
- ☐ If paper: Name of conference, date of conference, and place of conference
- ☐ If book chapter: Title of book, page range, publisher name and location
- ☐ If book: Publisher name and location
- ☐ If dissertation: Name of institution, type of degree, and department granting degree

Peabody Journal of Education, 93, 2

DOI or URL to published work (if available) https://www.tandfonline.com/doi/full/10.1080/0161956X.2018.1

Acknowledgement of Funding— Grantees should check with their grant officer for the preferred wording to acknowledge funding. If the grant officer does not have a preference, grantees can use this suggested wording (adjust wording if multiple grants are to be acknowledged). Fill in Department of Education funding office, grant number, and name of grant recipient institution or organization.

"This work was supported by U.S. Department of Education **[Office name]** IInstitute for Education Sci through **[Grant number]** R305B0900 to **Institution]** University of C. The opinions expressed are those of the authors and do not represent views of the **[Office name]** Institute for Education Sci or the U.S. Department of Education.

Rethinking Universal Suspension for Severe Student Behavior

Rebecca Hinze-Pifer
Stanford University

and

Lauren Sartain
University of Chicago Consortium on School Research

Correspondence should be sent to Rebecca Hinze-Pifer, Center for Education Policy Analysis, Stanford University, 520 Galvez Mall, Stanford, CA 94305. Email: rhp@stanford.edu. The authors appreciate support for this work from IES-PIRT Grant #R305B090025.

Abstract

Driven by a combination of concern for historically high suspension rates and substantial disproportionalities in suspension use, a recent wave of education reforms encourages schools to reduce their use of suspensions for student behavior management. Both academic and political discourse has focused on the extensive use of suspension for relatively minor behavioral infractions, with an implicit assumption or explicit articulation that suspension could still be used for severe infractions. This paper tests that assumption, providing evidence that reductions in suspensions for severe infractions may produce positive impacts without harming school safety.

Using data from high schools in the Chicago Public Schools (CPS), 2007-2014, we examine how declines in school reliance on suspensions for severe infractions are associated with changes in academic performance, attendance, and student reports of school climate for all students in the school. Recognizing the substantial methodological difficulty in obtaining impact estimates, we exploit a series of official and unofficial policy-induced changes to suspension practice, using school and student fixed effects models with extensive controls to reduce potential sources of bias in the estimates.

We find the reduction in out-of-school suspension for severe infractions was associated with small but statistically significant increases in student test scores, consequential attendance improvements (beyond the impact of fewer days suspended), and heterogeneity in changes to students' perceptions of school safety. Test score impacts are concentrated in racially diverse schools and those with low baseline suspension use. Attendance impacts are driven by schools predominantly serving African American students (which also had the highest baseline suspension rates); these schools also had large, significant improvements in perceptions of school climate.

Introduction

Local and national education leaders have broadly encouraged schools to reduce the use of exclusionary discipline like suspensions (e.g., U.S. Department of Education, 2014). Following national trends, the Chicago Public Schools (CPS) adopted a series of policy changes encouraging reduced use of exclusionary discipline between 2009 and 2014. In addition, district administrators used a variety of informal approaches to encourage school leaders to change school discipline practice. Schools reduced the share of high school students receiving an out-of-school suspension to 16% in 2013-14 from 24% in 2009-2010 (Stevens, Sartain, Allensworth, & Levenstein, 2015). The decline in suspension rates was not completely driven by schools changing their responses to low-level misbehavior—the use of suspensions has declined, even for severe infractions.

Drawing on incident-level infraction and discipline data combined with student-level administrative and survey data, this paper examines the relationship between changes in school suspension practice and changes in student test scores, grades, attendance, perceptions of safety, and perceptions of school climate. We focus on severe infractions committed by high school students.[1] Suspension practice for severe infractions is particularly salient because these infractions are frequently cited as almost universally suspension-worthy, and they account for more than one quarter of observed suspensions in our data.

The attempt to estimate causal impacts of suspension practice is fraught with a variety of methodological and conceptual difficulties. While there are clear correlations between suspensions and future outcomes, for both suspended students and for non-suspended students at

[1] Our definition of severe infraction is narrow—those the CPS Student Code of Conduct places in the three most severe categories. For example, two-person fights without injury are not severe. See the data section for a discussion of this choice and Table 1 for the included infractions.

schools with large numbers of suspensions (e.g., Balfanz, Byrnes & Fox, 2015), it is difficult to causally distinguish between the impacts of factors which cause misbehavior, the direct impacts of misbehavior, and the impacts of school use of exclusionary discipline in response to those behaviors. In addition, administrative discipline data is typically selected on multiple salient dimensions. Finally, researcher choices regarding the parameterization of school discipline practice and the population of interest may consequentially influence study conclusions.

During a period of plausibly exogenous policy-supported change in suspending behavior, we use school and student fixed effects models to examine how year-to-year changes in the use of out-of-school suspension (OSS) for severe infractions are associated with changes in academic performance, attendance, and student reports of school climate. Conceptualizing overall suspension practice as a school or grade-level treatment, we explore outcomes at the overall school level, as well as separately by student characteristics and by school types. This approach intentionally pools the impacts on suspended students, those at risk of suspension, and students who are not at risk of suspension, recognizing suspension practice may impact all students at the school, regardless of behavior or suspension status.

Background

Trends in the use of suspension in schools have largely been driven by rhetoric and theoretical arguments rather than empirical evaluation of the impact of suspension (see Steinberg & Lacoe, 2017 for a discussion). Identifying the causal impact of suspension policy is problematic on multiple dimensions. In particular, the types of behaviors that cause students to be suspended are themselves likely to impact student learning, requiring an identification strategy that separates the prevalence of suspension from the prevalence of disruptive behaviors. Further, suspension risk factors are strongly associated with the outcomes of interest, and school

officials are likely influenced by private information about students when making the choice to suspend. While large administrative data sets have become available in recent years, the discipline data generally suffer from consequential data quality problems, including underreporting of low-level events and inconsistent application of infraction definitions between schools and over time.

Little to no evidence suggests that widespread suspension use deters misbehavior or improves safety. In fact, after controlling for school and community factors, a disciplinarian approach seems to be associated with increased fear of crime at school (Schrek & Miller, 2003; Steinberg, Allensworth, & Johnson, 2011) and with higher teacher attrition (Mendez, Knoff, Ferron., 2002). In schools using suspension for minor infractions, students miss more school due to suspension without evidence the practice influences the prevalence of these behaviors (Morris, 2012; Skiba, 2000). In addition, there is some evidence that being suspended contributes to individual student anger and disconnection from school (Costenbader & Markson, 1998; Davis & Jordan, 1994; Jenkins, 1997).

A substantial body of descriptive work documents large correlations between suspension and a variety of negative future outcomes, both for suspended students and for students who are never suspended but attend schools with high suspension rates (Balfanz, Byrnes, & Fox, 2015; Davis & Jordan, 1994; Dawson, 1991; Fabelo et al., 2011). Furthermore, numerous recent studies document substantial inequality in the ways suspension policy differentially impacts students by race/ethnicity, gender, LGBT orientation, special needs, home language, and socioeconomic status (Fabelo et al., 2011; Himmelstein & Brückner, 2011; Losen & Gillepsie, 2012; Losen & Martinez, 2013; Losen, Hewitt, & Toldson, 2014; Osher et al., 2010; Porowski, O'Conner, & Aikaterini, 2014; Stevens et al., 2015).

Few studies have attempted to identify a causal link between suspension policy and student or school outcomes. Arcia (2006) compares matched suspended students and non-suspended students, finding large negative impacts of suspension on the affected student; however, suspended students had consequentially different pre-suspension trajectories from non-suspended students, suggesting matching may not have fully addressed the unobserved selection into suspension. Using data from a large urban district in Kentucky, Perry and Morris (2014) adopt a school fixed effects strategy to identify a plausibly causal estimate of the impact of suspension use on student test scores. They limit their analysis to students who are never suspended and find negative impacts of widespread suspension across school types. However, they parameterize school discipline policy as the share of students who are ever suspended, making it difficult to separate the impact of suspension from the impact of the events leading to suspension, even with controls for the number of events. Using Chicago data and a district policy change reducing the allowed length of suspension, Mader, Sartain, and Steinberg (2016) find shorter suspensions increase school attendance and test scores for impacted students, without negative spillovers on uninvolved peers. While also set in Chicago, our paper explores school practice along a different margin (the choice to suspend at all) – one which is relevant to a larger range of behavioral infractions. This paper aims to examine the associations between school suspension practice and multiple student outcomes in a way that simultaneously minimizes bias and appropriately parameterizes school-level suspension practice.

Changes to Suspension Policy in CPS

Beginning in the 2009-2010 school year, CPS initiated a series of reforms to school behavior policy, including substantial changes to the conditions under which school administrators could suspend (see Stevens et al. (2015) and Sartain et al. (2015) for detailed

descriptions of CPS discipline practice during the study period. In addition to the policy changes, district leaders consistently encouraged all schools to reduce their use of exclusionary practices. Instead, schools were encouraged to adopt a variety of restorative practices, including community service, restorative group conferencing, victim impact panels, and student-run disciplinary proceedings. This climate in the district targeted the full spectrum of punishment actions, including reduced suspension length, substitution of in-school suspension (ISS) for out-of-school suspension (OSS), and support for consequences which kept students in the classroom as much as possible. Figure 1 documents the general reduction in the percentage of students suspended and in the average number of days suspended, as well as the partial substitution of ISS for OSS.

(Figure 1 about here)

Culture of Calm, 2009-2011

In 2009-2010, CPS piloted an initiative called the Culture of Calm (CC) in six high schools, selected due to a combination of student trauma exposure and high suspension use. Programming was designed to provide students with the individualized supports they needed to develop social-emotional skills, aiming to change the way students approached situations of tension and conflict. Schools were also trained in the use of restorative justice programs in lieu of suspension. The following year, CPS devoted $40 million of the money it received under the American Recovery and Reinvestment Act (ARRA) to expand the implementation of CC programming into an additional 40 high schools. District officials also encouraged non-CC schools to adopt similar reforms, albeit without financial support. After ARRA funds expired, CPS devoted ongoing resources to supporting schools in switching away from suspensions to consequences rooted in restorative justice within a climate of social-emotional supports.

Changes to the Student Code of Conduct, 2011-2014

The Student Code of Conduct (SCC) provides formal definitions of proscribed behaviors in CPS schools and the actions school leaders may take in response to each behavior. It is adopted at the district level but allows substantial discretion in how school leaders handle most infraction types. The SCC is revised each summer, generating policy-induced variation in the use of exclusionary practices after major revisions. Revisions to the SCC during this period did not change the severity level of any behavior, although they did occasionally add proscribed behaviors or clarify the definition of a behavior. During the study period, revisions primarily changed the set of actions available to schools in response to specific infractions.

In response to a state mandate to implement a formal Response-to-Intervention model for school behavioral support, and in order to provide guidance regarding the discipline practices encouraged by the CC program, CPS revised the SCC in the summer of 2011. These revisions clarified the restorative practices endorsed by the district and drastically expanded the range of behaviors for which restorative practices were recommended. In the summer of 2012, CPS further revised the student code of conduct to move towards more restorative/non-exclusionary approaches, making it more difficult to assign suspensions longer than five days, eliminating mandatory ten-day suspensions for the most severe infractions, and reducing suspension length for low-level infractions (see Mader, Sartain, & Steinberg, 2016 for an analysis of the impact of the change to long suspensions). In February 2014, CPS announced its Suspension and Expulsion Reduction Plan, formally codifying some of the changes in practice it had been informally encouraging. In addition, it outlined a set of changes that would be made to the SCC in the summer of 2014 and made changes to the expulsion process.

Data

In this paper, we use student-level CPS administrative data from the 2007-08 to 2013-14 school years, as well as My Voice, My School (MVMS) survey data administered by the UChicago Consortium to all high school teachers and students in spring 2009 and 2011-2014. The sample includes all traditional high school students in public schools with enrollments over 30 students; it does not include students in charter[2] or selective enrollment schools.

CPS administrative data include demographic and background information, including race/ethnicity, gender, special education status, birth date, and students' residential Census block group. We include a poverty index based on block-group median income and male employment and a social status index based on block-group mean education and professional employment, as well as a residential mobility measure. We also know if a student has a substantiated history of abuse or neglect as reported in Illinois Department of Child and Family Services (DCFS) data. CPS enrollment information includes the school the student attends, grade level, and both days enrolled and days in attendance.

We use disciplinary infraction and suspension data to construct treatment variables to describe suspension practices at the school-year level, described in additional detail below. For each school-reported infraction a student commits, the discipline data include which portion(s) of the student code of conduct was violated, the date of the offense, and suspension information.

The CPS SCC categorizes disciplinary offenses as levels 1 through 6, where level 1 is extremely minor and level 6 is extremely serious and illegal (outlined in Table 1). We characterize an event as severe if it included a level 4 or higher infraction—the point at which most infractions are nominally illegal. Throughout this study, we assume that low-level events

[2] Charter schools are not required to report student-level discipline information to the district.

are underreported and that there may be time trends in reporting levels. We are relatively confident that nearly all severe events are reported whenever school officials are aware of them and that events leading to a suspension are present in our data, regardless of level.

(Table 1 about here)

In order to control for secular trends in student behaviors (i.e., students are misbehaving less or more over time), we use a combination of administrative behavior data and survey-based measures of behavioral disruption from teachers and students (UChicago Consortium on School Research, 2017). All MVMS survey data are standardized using the mean and standard deviation of the 2008-09 data. This anchoring of the survey measures in the first year of outcomes data allows us to see schools improving over time relative to other schools in the district. Prior work on this topic generally did not include student and teacher-reported controls for behavior. In our data, these measures are predictive of both treatment and outcome even after controlling for administrative reports of student behavior.

CPS administrative data include a variety of academic outcomes at the student level, which we examine in this paper. During the period of the study, CPS used the ACT's Educational Planning and Assessment System (EPAS) testing program in high schools with testing points in grades 9-11. In order to account for variations in the test year-to-year, we standardize the test scores within grade level and school year. We also include a predicted measure of eighth-grade achievement prior to the students entering high school. That measure uses a student's entire test score trajectory from grades 3-8 on the state standardized test in Illinois to predict academic ability at the end of grade 8. We also look at grade point average

(GPA) in high school, as well as the proportion of enrolled days students are in attendance, net of days suspended.[3] Finally, we consider MVMS measures of overall school climate.

Analytic Sample

School suspension practice likely impacts all students in a school. While we expect it to be most salient to students actually suspended or directly at risk of suspension due to behavior, school suspension practice is observed by all students in a school. In schools where suspension is widespread, most students will have friends who are suspended and may perceive themselves as at high risk of being suspended (see Ferguson, 2000 and Nolan, 2011). Furthermore, one core argument in favor of suspension use—particularly for severe infractions—suggests it should produce positive impacts for uninvolved students by preventing disruptions and removing disruptive students (see Curran, 2016 for a discussion). Accordingly, we include all of a school's students in the analytic sample, estimating a net impact across all students.

Table 2 describes students in the analytic sample (see the next section for a description of the suspension risk groupings), with each student appearing once for each year they are enrolled in CPS. We provide descriptive statistics for all students in the sample, as well as students who are at various levels of risk for suspension. Overall, the analytic sample is 46 percent black and 43 percent Latinx, as well as 51 percent male. Fifteen percent of the sample has been identified for special education services, most qualify for reduced- or free-price lunch (85%), and eight percent have a substantiated history of child abuse or neglect. Prior to entry into the sample, the average student had 2.1 residential moves and 2.4 school switches (this includes the transition to high school for almost all students).

---

[3] Student attendance is calculated as the number of days present in school, divided by the difference between the number of days the student was enrolled and the number of days suspended.

(Table 2 about here)

More than one in five students in the sample received at least one out-of-school suspension (OSS) in any given year, and 9% committed a severe infraction. Some 14% of students received an in-school suspension (ISS) in any given year; the majority of these students also received an OSS in that year. The average student experienced ISS for 0.46 days during the year and OSS for 1.40 days.

Predicting Suspension Risk

There may be variation in estimated effects of changes in school suspension practices by a student's likelihood to be suspended. For example, we might expect that students highly likely to experience suspension would be the most impacted by a reduction in school suspensions. Given the panel structure of our data, we are able to explicitly model the extent to which the full range of available student characteristics (including prior year disciplinary record, attendance, test scores, and grades) predict their propensity to be suspended in any given year. In order to define suspension risk in a stable way, we use student characteristics from the 2007-08 school year and suspension data from 2008-09 to model the patterns of suspension in the baseline year. The resulting measure thereby captures students' revealed risk based on actually observed suspension practices in Chicago schools. We think it is important to use data prior to the year where the district started emphasizing reductions in suspension, as one's propensity to be suspended for the same behavior will have likely declined post-policy changes. This allows us to estimate the propensity that a student in 2013-14, for example, would have experienced suspension in earlier years. Specifically, we use the coefficients from the model to produce an estimated propensity the student would have been suspended under the 2008-09 disciplinary

regime for each subsequent year based on the previous year's data. We use a similar approach to estimate each student's risk of having a severe infraction.

Table 2 characterizes students with three different levels of predicted suspension risk where the low-risk group is roughly the bottom half of students, with an estimated propensity at or below 0.2, and the moderate-risk group has an estimated propensity between 0.2 and 0.5. The high-risk group has an estimated propensity above 0.5 and comprises approximately the top risk decile. Consistent with prior research on the correlates of suspension, students in the high-risk group are disproportionately black, male, and low income. They are also a more mobile student population. Further, as we would expect given the construction of the risk groups, students we estimate with high likelihoods of suspension are indeed more likely to commit severe infractions and experience both in-school and out-of-school suspension.

Analytic Framework

Between 2008 and 2013, the percentage of high school students in CPS receiving OSS fell from 23% to 16%, and the share of severe infractions receiving OSS fell from 93% to 84%. Over the same time period, the prevalence of severe behavior incidents remained statistically constant at roughly 12 incidents per 100 students. Given the constant time trend in severe events, we contend the decline in suspensions arose from largely exogenous forces including district policy and shifting social norms surrounding the use of suspension. Due to the informal nature of the policy changes, changes to suspension practice were adopted with significant temporal variation at the school level.

Accordingly, we begin with a school fixed effects model to explore impacts of reducing suspension use. School fixed effects account for all time invariant school-level characteristics; however, they do not address the reality that many important aspects of schools change from

year to year. In particular, as the population of a school evolves with each new cohort, student characteristics change, including both factors likely to impact outcome variables and factors associated with problematic behavior. We also include fixed effects for the year and student grade level in all models, flexibly accounting for mean differences between years and grade levels. Our preferred specification adds fixed effects at the student level, accounting for all time-invariant student-level factors, combined with extensive controls for temporal changes at the student and school levels. This approach functionally uses individual students as controls for themselves—comparing their outcomes in years with higher suspension use to their outcomes in years with lower suspension use, after controlling for observed time-varying student characteristics.

The identifying assumption of fixed effects models is that there are not unobserved time-varying factors which systematically impact both changes in the school-level use of suspensions and outcomes for students at that school. To the extent year-to-year changes in school use of suspension is entirely exogenous, this is certainly the case. However, if there are meaningful year-to-year differences in student behavior, we would expect these to impact both outcomes and suspensions. For example, if a school has undergone efforts to increase student engagement, those efforts could simultaneously improve learning and decrease suspensions by decreasing misbehavior. We address this concern in two ways. First, all models include controls for student and teacher perceptions of safety and student behavior at the school, as well as counts of the per-capita number of infractions at each severity level (results are also robust to controls for the prevalence of infractions by type). Few prior papers include controls for administrative discipline data, and we are unaware of any prior work including student or teacher reports of school safety or behavioral disruption. The addition of these controls is important, as the survey-based

measures predict treatment (regardless of how treatment is defined) even when we control for administrative reports of behavior and school fixed effects. Second, we define our treatment variable (described in the next section) in a way that we argue is likely uncorrelated with the level of behavioral disruption in the school.

We present three increasingly conservative models. The base model is of the form:

$$Y_{ist} = \alpha_1 + \alpha_2 D_{st} + X'_{st}\beta_1 + \eta_s + \gamma_g + \pi_t + \varepsilon_{ist} \quad (1)$$

Y_{ist} is the outcome in the current year for student i attending school s in year t. D_{st} is a measure of suspension practice in that school and year (described in the next section), and X_{st} is the set of time-varying school characteristics. $\eta_s, \gamma_g,$ and π_t are school, grade-level, and year fixed effects. Model (1) includes controls for enrollment and school-level averages of student demographic characteristics (gender, race/ethnicity, free/reduced price lunch, special education, student census block poverty, student abuse/neglect history, and prior retention).

Model (2) adds X_{ist}, an extensive set of student-level controls. These include demographic characteristics (gender, race/ethnicity, free/reduced price lunch, special education category, student census block poverty, student abuse/neglect history, and prior retention, school, and residential mobility history), prior standardized test scores, and lagged student disciplinary history (flags and counts of administrative reports at each of the six incident levels).

$$Y_{ist} = \alpha_1 + \alpha_2 D_{st} + X'_{st}\beta_1 + X_{ist}\beta_2 + \eta_s + \gamma_g + \pi_t + \varepsilon_{ist} \quad (2)$$

Our preferred specification, model (3), replaces the lagged outcome variables with student fixed effects, ϕ_i. The addition of student fixed effects fully controls for all time invariant student-level characteristics. Recognizing adolescents change in important individual ways during high school, we still include time-variant demographic and lagged disciplinary controls,

$\tilde{X}_{ist}$. Results tables also include a second set of results from this model, where D_{st} is determined at the grade rather than school level.

$$Y_{ist} = \alpha_1 + \alpha_2 D_{st} + X'_{st}\beta_1 + \tilde{X}_{ist}\beta_2 + \phi_i + \eta_s + \gamma_g + \pi_t + \varepsilon_{ist} \quad (3)$$

All tables report standard errors clustered at the school level; results are robust to alternate clustering approaches. Due to the degrees of freedom adjustment required for the standard errors if students switch schools,[4] we drop students who switch schools, causing the η_s school fixed effect in model (3) to be absorbed into ϕ_i (point estimates are substantively unchanged).

Defining School Suspension Practice

The fundamental problem in defining treatment (i.e., D_{st}, the school discipline policy) is that suspensions are school reactions to student behavior. Some factors which influence behavior certainly also influence outcomes at the individual level, and peer behavior likely influences most outcomes as well. While we include extensive controls, the primary threat to our approach remains the possibility that some factor simultaneously influences the treatment and the outcome. One way we address this limitation is by selecting a treatment variable which captures the extent to which student misbehavior is addressed via suspensions, but which is relatively uncorrelated to the base level of disciplinary disruption present in the school.

We parameterize school suspension practice as the percentage of severe events (levels 4-6) which led to an out-of-school suspension (OSS). The trend in school use of OSS to respond to severe events is shown in Figure 2. In 2008-09, schools gave students who committed severe

[4] For student fixed effects models, the cluster-robust estimator requires an inefficient degrees of freedom adjustment when students are not nested in schools. The magnitude of results do not appreciably change with mobile students included, but the adjusted standard errors are larger.

infractions OSS nearly 95 percent of the time. By the end of the period, 85 percent of severe infractions resulted in OSS. In other words, in 2013 at a typical high school, 15 percent of the time a student committed a very high-level offense, the school responded in a way that did not involve removing the student from the building.[5] Higher values of the treatment variable may be interpreted as reflecting school practice which tends to use suspension more readily. We exclusively focus on OSS because the policy guidance given by CPS during the study period encouraged schools to use in-school suspension (ISS) as a partial substitute for OSS, and the nature of ISS varies substantially across schools. In particular, some schools use ISS time to implement programming of a restorative or rehabilitative nature. Furthermore, it is entirely plausible that ISS may be substantively different than OSS (even if unprogrammed), insofar as it may help keep students engaged with school.

This approach to measuring school suspension is grounded in the understanding that school administrators establish suspension practice by the series of choices they make in response to specific behaviors—it captures an extent to which the discipline regime is punitive, conditional on the antecedent behavior. For example, some administrators may universally suspend students involved in minor fights, while others may be more selective about which fights result in suspensions. See the Testing for the Reflection Problem section below for a discussion on whether this variable dynamically responds to student behavior. This measure is relatively uncorrelated with the level of underlying behavioral disruption at a school (detailed results available on request), meaning schools which readily suspend students for severe infractions have varying levels of underlying behavioral disruption. Furthermore, if we limit our

[5] Expulsions are rare, and trending flat or down during this period. These numbers are substantively unchanged when infractions leading to expulsion proceedings are removed – largely because those infractions typically involve a suspension prior to the expulsion hearing.

measures to relatively common infractions expected to be well-reported in our data, there is a strong within-school, within-year correlation between the share of reported infractions resulting in suspension in each severity class. While we do not directly observe the school decision-making process about reporting student misconduct, we suspect this correlation arises from the broad reality that a small number of people in a school are responsible for making most suspension decisions (principals, assistant principals, and/or deans of discipline).

(Figure 2 about here)

We prefer measuring suspension practice as the share of severe events leading to OSS for several reasons. First, we suspect, and school officials affirm, that when severe behavioral infractions come to the attention of school personnel, they are nearly universally reported in the administrative data, suggesting we are able to measure this variable with relatively little error. Second, it shows a similar downward time trend to other plausible treatment variables, suggesting suspensions for severe infractions responded to the exogenous pressure from the school district to reduce overall suspensions. In addition, the existing literature offers little insight regarding the use of suspension for severe infractions. Finally, unlike other measures of suspension policy, it is only weakly correlated with measures of student behavior, which we argue make it less susceptible to omitted variables bias.

Testing for the reflection problem

Our approach is threatened by an additional, more subtle concern. It is plausible that suspension policy itself influences the prevalence of student behavior (either positively or negatively). We estimate a net impact, which collapses the impacts of student behavior change, the impact of suspension (or not) on the potentially suspended, and any spillover effects on uninvolved students. However, this estimation approach can produce a biased estimate of the net

impact if the treatment is simultaneously caused by student behavior (Manski, 1993). Of particular concern is the possibility that school officials may determine suspension policy dynamically, in response to recent behavior (e.g., "cracking down" after a series of severe incidents).

Using infraction-level data including the date of each infraction, we searched for evidence school administrators were behaving in this way. We performed a series of analyses to determine whether our treatment variable (the share of severe infractions receiving OSS) is predicted by recent behavioral infractions and failed to find evidence of the refection problem. Table 3 displays the results of one such analysis, where we regress the treatment variable measured over a two-week period, t, on the per-student number of infractions in each of four prior periods. Results are substantively similar when we restrict the sample to schools with severe events in most weeks, suggesting the insignificant coefficients are not driven by noise in measurement of the treatment variable. They are also unchanged by varying the temporal period between 1 and 4 weeks or by measuring lagged infractions by the nature of the infraction. In aggregate, we do not find evidence school administrators are adjusting their use of suspensions in response to recent student behavior.

(Table 3 about here)

Results

Throughout this section, we focus on the estimates produced by the student fixed effects model with treatment defined at the school level, using equation 3. Table 4 presents estimation results for three outcomes measures for the various models discussed above. Estimates in column (1) are from the school fixed effects model with time-variant school-level controls, while those in column (2) include extensive student-level controls and those in column (3) include student

fixed effects and time-varying student controls. Recognizing disciplinary choices may be determined at the grade level rather than the school level, column (4) uses the same estimation approach as column (3), but measures suspension practice at the grade level rather than the school level. Negative coefficients imply reducing suspensions is consistent with positive consequences. For discussion purposes, we divide the coefficients by 10—reflecting a 10 percentage point change in school suspension policy, which is roughly representative of the extent of change observed within the typical school during the study period.

In aggregate, we find reducing suspensions is associated with a small significant positive change in student test scores, which is robust to specification. Separate results for math and reading produce similar patterns and conclusions; the point estimates are larger for math than reading, but without a statistical difference between the values. Our preferred specification suggests a 10 percentage point decline in suspension for severe events is associated with a 0.022σ improvement in test scores.

Reducing suspension for severe infractions by 10 percentage points is associated with a 0.5 percentage point increase in the share of days students attend school (approximately one day annually per student). The attendance variable has been adjusted for the fact that suspension mechanically produces absences; this increase is above the automatic increase in attendance produced by fewer out-of-school suspensions.

(Table 4 about here)

Examining student grades, none of the specifications provide any evidence changing suspension policy impacts overall student GPA (results available on request). Perhaps more importantly, the last panel in Table 4 shows no indication students overall experience schools as less safe when suspension use for severe infractions declines. The measure of school safety is

constructed using Rasch analysis (Luppescu & Ehrlich, 2012) to combine student responses on a 4-point Likert scale rating how safe they feel in the hallways and bathrooms, outside around school, traveling between home and school, and in class (see UChicago Consortium on School Research, 2017 for detailed survey documentation). These results have large standard errors, so we do not claim they obviate the concern that reducing suspensions may reduce safety. However, the direction is consistent with *increased* perceptions of safety with fewer suspensions. We also separately tested the likelihood a student would endorse each of the safety questions using a logit model, with consistently small, insignificant results in a direction consistent with increased safety. Estimates of impacts on other aspects of school climate were also small and insignificant.

Heterogeneity analysis

It is reasonable to expect school discipline regimes will have heterogeneous impacts by student characteristics and school types. At the school level, we examined results separately for schools with high baseline suspension rates compared with schools with moderate or low rates (based on per capita suspensions in 2008-09). We also estimated separate treatment effects by school racial composition, where schools are identified as having mostly African American students (>75%), mostly Latinx students (>75%), or being racially diverse.[6]

(Table 5 about here)

When impacts are separately estimated for relevant subgroups of schools, we find evidence consistent with differential impacts by school racial composition for Chicago schools. Table 5 presents separate estimates by school baseline suspension rates (top panel) and school

[6] This approach to grouping Chicago schools is used in prior UChicago Consortium analyses (e.g. Sartain, Allensworth & Porter, 2015). Students of all races are counted as Latinx if their ethnicity variable indicates such. The African American group includes all non-Latinx students whose administrative race variables indicates such. School race/ethnicity classification is robust to classifying African American Latinx students in the African American group instead.

racial composition (bottom panel). It should be noted that schools serving predominantly African American students had substantially higher baseline suspension rates than other schools. It is possible that the observed heterogeneity by school type is due to correlates of racial composition rather than racial composition itself. In schools with low baseline suspension use and in schools serving more racially diverse students, reducing the use of suspensions for severe infractions is consistent with improvements in standardized test scores.

Earlier analysis did not find any significant aggregate changes in school climate variables with changes in suspension for severe infractions, with imprecise estimates. However, large significant differences emerge when schools serving predominantly African American students are considered separately, as shown in Figure 3. Among schools with largely African American students, reducing suspensions is associated with increased reports that specific teachers supported student learning (teacher-student relationships), that students generally had good relationships with teachers (supportive teachers), and that students felt connected to their school (school connectedness).[7] The opposite pattern occurred for schools serving racially diverse or predominantly Latinx student populations. Scaling these results to reflect a typical change in suspension practice, student reports of school climate improve by $0.10\sigma - 0.13\sigma$ with a 10

[7] Students rate their agreement with the following statements on 4-point Likert scales, with measures constructed using Rasch analysis (UChicago Consortium on School Research, 2017).
Student-Teacher Relationships: When my teacher tell me not to do something, I know they have a good reason; I feel safe and comfortable with my teachers at this school; My teachers always keep their promises; My teachers always listen to students' ideas; My teachers treat me with respect.
Supportive Teachers: The teacher for this class… Helps me catch up if I am behind; Is willing to give extra help on schoolwork if I need it; Notices if I have trouble learning something; Gives me specific suggestions about how I can improve my work in this class; Explains things in a different way if I don't understand something in class.
School Connectedness: I feel like a real part of my school; People here notice when I'm good at something; Other students at my school take my opinions seriously; People at this school are friendly to me; I'm included in lots of activities at school.

percentage point decline in the use of suspension for severe infractions. Mostly African American schools also show evidence of improved attendance with reduced suspension use. After adjusting for days suspended, students attend 1.5% more of their enrolled days with a 10 percentage point decline in suspension practice. Troublingly, students at schools serving predominantly Latinx students report increased concerns about safety with reduced use of suspension for severe infractions. The results are substantively unchanged if we use other break points to define the baseline suspension categories, if we divide schools by baseline treatment or baseline disciplinary infractions, if we define the baseline suspension rates and racial share variables continuously, or if we use alternate model specifications.

(Figure 3 about here)

At the student level, we examine heterogeneity by student propensity to be suspended and propensity to have a severe infraction (see the data section for a description of the propensity score estimation). It should be noted that the high degree of racial segregation in Chicago Public Schools makes it difficult to separate the extent to which differences by race arise at the student or school level (or both), although it is interesting to note the school environment and attendance impacts are much larger when estimated for mostly African American schools than for African American students (results available on request). Table 6 suggests test score changes are larger for students who are at relatively low risk of having a severe infraction, while students at high risk of OSS experience schools as *safer* when suspension use for severe infractions declines.

(Table 6 about here)

Limitations

Our approach in this paper fundamentally relies on the assumption that informal encouragement to reduce suspensions and formal policy changes at the district level over a

period of five years induced exogenous changes in the use of suspension for severe behavioral infractions, allowing identification based on idiosyncratic variation in the year in which schools implemented the suggested reforms. The primary threat to this approach is the possibility that some time-varying characteristic at the school level is simultaneously correlated with treatment and the error term. While available anecdotal evidence does not suggest this was the case, it is possible that principals (or deans of discipline) who disproportionately responded to district pressure were also more likely to implement other types of programs that produced the observed impacts.

In interpreting the results by school types, we are unable to determine whether the observed patterns emerge due to differences in school racial composition, baseline suspension use, or some other correlated school-level factor. Similarly, we are unable to say whether the student-level heterogeneity by estimated risk is directly due to the risk they experience suspension or due to some contributing factors.

Finally, while the suspension measure we use is relatively uncorrelated with observed school use of suspension for less extreme infractions, data quality limitations prevent us from eliminating the possibility that our measure is a strong proxy for suspension use overall—in which case, these results should be interpreted as addressing suspension use overall, rather than strictly suspension for severe infraction. Similarly, there is some correlation between reductions in suspension use and reductions in the lengths of suspensions, and it is possible some portion of the measured impacts arise from adjustment along that margin.

Discussion

Recent focus on reducing the extensive use of suspension in response to low-level student misbehavior provides little insight regarding the circumstances under which suspension may be

an appropriate strategy. Informal conversations with school personnel suggest a narrow set of circumstances for which there is broad support for ongoing use of suspension – e.g., "cooling off" periods after severe infractions involving conflict or short-term removal after an extreme event while the school pursues expulsion or alternate placement. During an era of suspension reform, it is less clear precisely which other circumstances, if any, might be best addressed by suspension. Examining the data behind Figure 2 reveals that Chicago largely achieved the observed decline by shifting the severity threshold at which suspension is imposed (e.g., assigning ISS rather than OSS for less severe weapons, while still suspending for firearms, switchblades, and explosives), a strategy which seems a natural direction for reform.

By examining current suspension practice for severe infractions, this project interrogates the practice of near-universal suspension for extreme student behaviors. At the current margin of practice in Chicago, reducing suspension use for severe infractions seems consistent with a set of generally positive but small impacts. Insofar as student perceptions of safety accurately reflect actual safety, we do not find broad evidence reducing suspensions for severe infractions makes schools less safe; however, there was a decline in reported safety with suspension reduction at schools serving predominantly Latinx students, suggesting changes in practice should be pursued with cautious attention to the ongoing security concerns specific to each school context.

In determining the practical implications of this study, the counterfactual is crucially important; our estimates compare out-of-school suspension with an unknown set of alternate approaches. For severe infractions, schools almost certainly replaced out-of-school suspensions with other disciplinary actions. In our data, it seems approximately one quarter of the reduction in out-of-school suspension is replaced with in-school suspension. Anecdotal evidence suggests some schools were adopting restorative approaches for other infractions. Additional work is

needed to determine the relative impacts of these other approaches and whether our findings are truly due to the removal of suspension or arise from alternate programming students received after incidents.

Acknowledgements

The authors thank the Chicago Public Schools for ongoing data access and support. We also thank workshop and conference participants at the University of Chicago, the University of Illinois, Stanford University, the Association for Public Policy and Management, the Association for Education Finance and Policy, and the Society for Research on Educational Effectiveness, researchers at the University of Chicago Consortium on School Research, Stephen Raudenbush, Ofer Malamud, Amy Claessens, and Stephen Baker for helpful comments and discussions. Hinze-Pifer is grateful for support from the IES-PIRT fellowship (PR/Award #R305B140009), the University of Chicago Committee on Education, and the National Academy of Education/Spencer Dissertation Fellowship. Sartain gratefully acknowledges the Atlantic Philanthropies, as well as the Spencer Foundation and the Lewis-Sebring Family Foundation, whose operating grants support the work of UChicago Consortium.

References

Arcia, E. (2006). Achievement and enrollment status of suspended students: Outcomes in a large, multicultural school district. *Education and Urban Society, 38*(3), 359-369.

Balfanz, R., Byrnes, V. & Fox, J. (2015). Sent home and put off-track: The antecedents, disproportionalities, and consequences of being suspended in the ninth grade. *Journal of Applied Research on Children: Informing policy for children at risk, 5*(2), 13.

Chicago Public Schools. (2015). Student Code of Conduct. Retrieved from: http://cps.edu/Pages/StudentCodeofConduct.aspx.

Costenbader, V. & Markson, S. (1998). School suspension: A study with secondary school students. *Journal of School Psychology, 36*(1), 59-82.

Curran, F. C. (2016). Estimating the effect of state zero tolerance laws on exclusionary discipline, racial discipline gaps, and student behavior. *Educational Evaluation and Policy Analysis, 38*(4), 647-668.

Davis, J. E. & Jordan, W. J. (1994). The effects of school context, structure, and experiences on African American males in middle and high schools. J*ournal of Negro Education, 63*(4), 570-587.

Dawson, D. A. (1991). Family structure and children's health and well-being: Data from the 1988 National Health Interview Survey on Child Health. *Journal of Marriage and Family, 53*(3), 573-584.

Fabelo, T., Thompson, M.D., Plotkin, M., Carmichael, D., Marchbanks, M.P. III, & Booth, E.A. (2011). *Breaking schools' rules: A statewide study of how school discipline relates to students' success and juvenile justice involvement.* New York, NY: Council of State Governments Justice Center.

Ferguson, A. A. (2000). *Bad Boys: Public Schools in the Making of Black Masculinity*. Ann Arbor: University of Michigan Press.

Himmelstein, K. E. W., & Brückner, H. (2011). Criminal justice and school sanctions against nonheterosexual youth: A national longitudinal study. *Pediatrics, 127*(1), 49-57.

Jenkins, P. H. (1997). School delinquency and the school social bond. *Journal of Research in Crime and Delinquency, 34*(3), 443-472.

Losen, D.J., & Gillespie, J. (2012). *Opportunities suspended: The disparate impact of disciplinary exclusion from school.* Los Angeles, CA: The Civil Rights Project/Proyecto Derechoes Civiles.

Losen, D.J., & Martinez, T.E. (2013). *Out of school and off track: The overuse of suspensions in American middle and high schools.* Los Angeles, CA: Civil Rights Project/Proyecto Derechoes Civiles.

Losen, D.J., Hewitt, D., & Toldson, I. (2014). *Eliminating excessive and unfair exclusionary discipline in schools: Policy recommendations for reducing disparities.* Bloomington, IN: The Equity Project at Indiana University.

Luppescu, S. & Ehrlich, S. (2012). *The new and improved CCSR Rasch Measurement Model Primer*. Retrieved from http://consortium.uchicago.edu/sites/default/files/uploads/survey/Overview%20of%20Analyses_0.pdf.

Mader, N., Sartain, L., & Steinberg M. P. (2016, March.) *When suspensions are shorter: The effects on school climate and learning.* Paper presented at the Association for Education Finance & Policy's Annual Conference, Denver.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies, 60*(3), 531-542.

Mendez, L. M. R., Knoff, H. M., & Ferron, J. M. (2002). School demographic variables and out-of-school suspension rates: A quantitative and qualitative analysis of a large, ethnically diverse school district. *Psychology in the Schools, 39*(3), 259-277.

Morris, M. W. (2012). *Race, Gender, and the School to Prison Pipeline: Expanding our Discussion to Include Black Girls*. New York: African American Policy Forum.

Nolan, K. (2011). *Police in the Hallways: Discipline in an Urban School.* Minneapolis, MN: University of Minnesota Press.

Osher, D., Bear, G.G., Sprague, J.R., & Doyle, W. (2010). How can we improve school discipline? *Educational Researcher, 39*(1), 48-58.

Perry, B. L. & Morris, E. W. (2014). Suspending Progress: Collateral Consequences of Exclusionary Punishment in Public Schools. *American Sociological Review, 79*(6), 1-21.

Porowski, A., O'Conner, R., & Aikaterini, P. (2014). *Disproportionality in school discipline: An assessment of trends in Maryland, 2009-12.* Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education and Regional Assistance, Regional Educational Laboratory Mid-Atlantic.

Sartain, L., Allensworth, E. M., & Porter, S. (2015). *Suspending Chicago's Students: Differences in Discipline Practices across Schools.* Chicago, IL: University of Chicago Consortium on Chicago School Research.

Schreck, C.J., & Miller, J.M. (2003). Sources of fear of crime at school: What is the relative contribution of disorder, individual characteristics, and school security. *Journal of School Violence, 2*(4), 57–79.

Skiba, R. J. (2000). *Zero Tolerance, Zero Evidence: An Analysis of School Disciplinary Practice.* Bloomington, IN: Education Policy Center.

Steinberg, M.P., Allensworth, E., & Johnson, D.W. (2011). *Student and teacher safety in Chicago Public Schools: The roles of community context and school social organization.* Chicago, IL: University of Chicago Consortium on Chicago School Research.

Steinberg, M. & Lacoe, J. (2017). What do we know about school discipline reform? Assessing the alternatives to suspensions and expulsions. *Education Next, 17*(1), 44-52.

Stevens, W. D., Sartain, L., Allensworth, E.M., & Levenstein, R. (2015). *Discipline Practices in Chicago Schools: Trends in the Use of Suspensions and Arrests.* Chicago, IL: University of Chicago Consortium on Chicago School Research.

UChicago Consortium on School Research. (2017). *Survey documentation.* Retrieved from: https://consortium.uchicago.edu/surveys/documentation.

U.S. Department of Education. (2014). *Rethinking School Discipline*. Available from: http://www2.ed.gov/policy/gen/guid/school-discipline/fedefforts.html#guidance

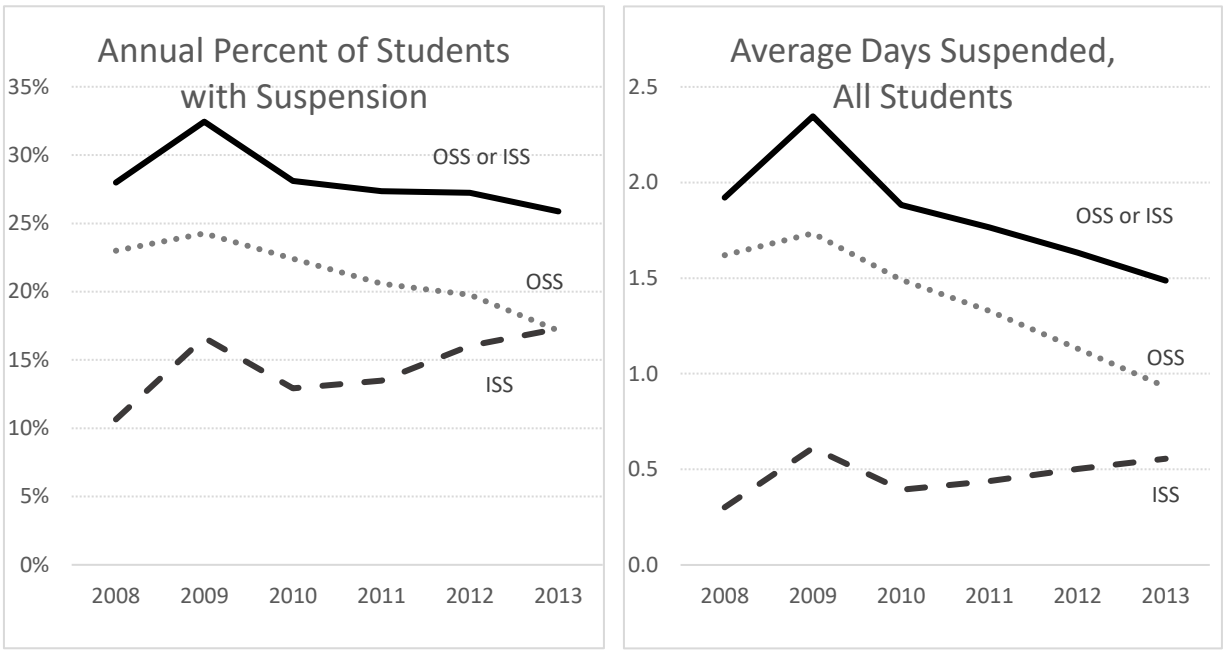

Figure 1: Time trends in suspensions, by suspension type.
Notes: All high school students in traditional CPS schools. Year is the fall of the school year.

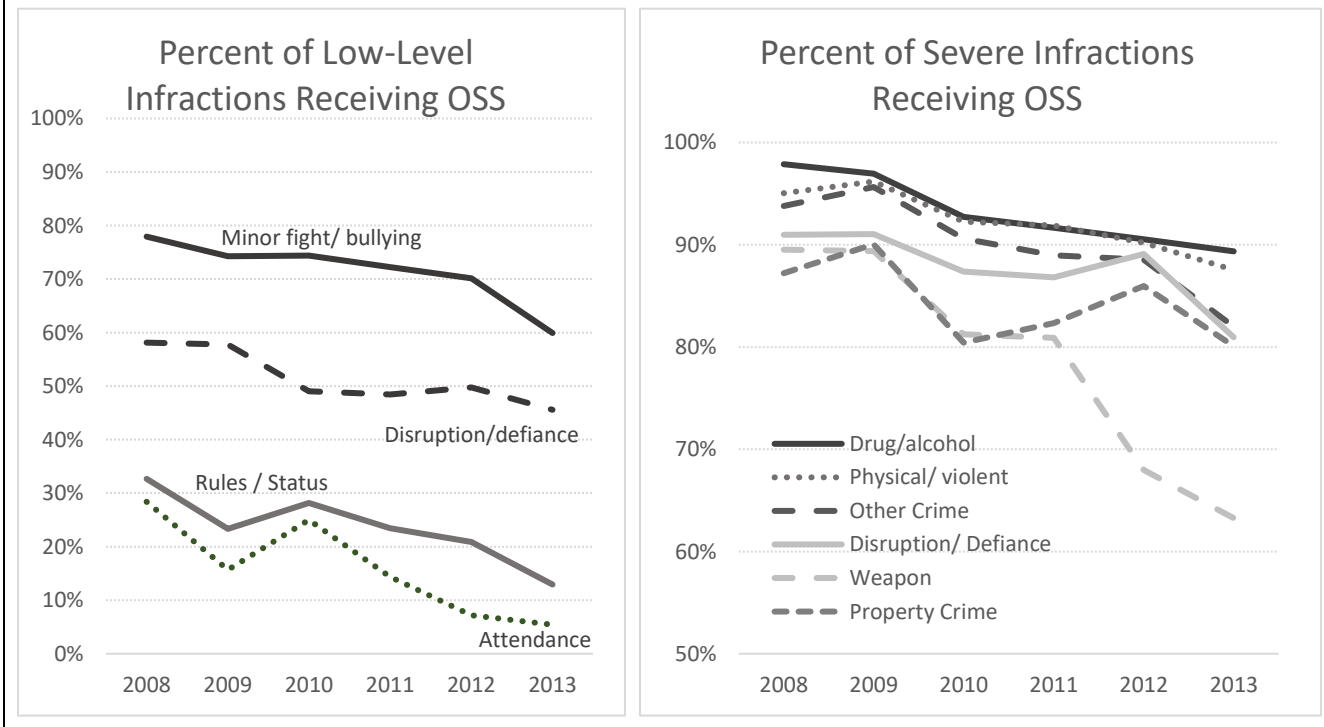

Figure 2: Trend in share of infractions resulting in an out-of-school suspension, by infraction type.
Notes: Scale on two panels is different for readability. Year is the fall of the school year.

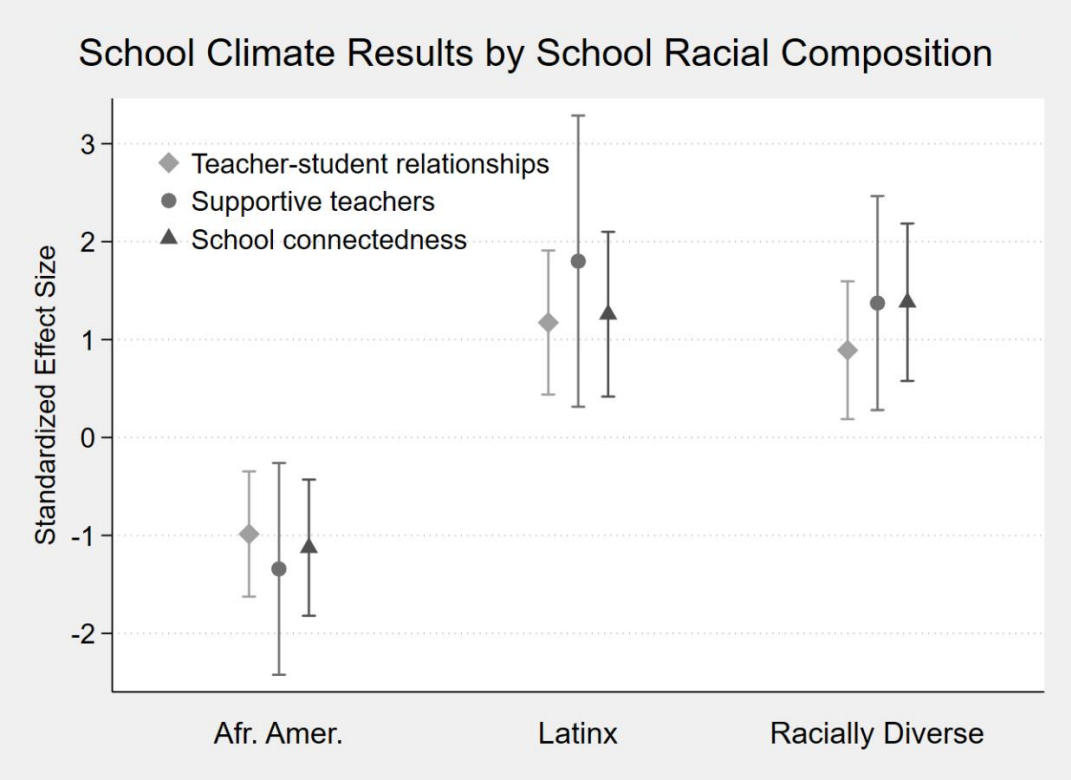

Figure 3: School climate results by student population race/ethnicity groupings

Notes: Point estimates are regression coefficients from model (3). Standard errors are clustered at the school level. Schools serving predominantly African American or Latinx students have 75% of their student population in those groups. Negative values imply improved climate when suspension is reduced.

Table 1: CPS Code of Conduct Disciplinary Infractions

		Infraction
Lower Level	1	Running/noise in halls, leaving or skipping class, persistent tardiness, disrupting class, loitering, unsanctioned computer use, cell phones
	2	Posting bills, leaving school, walkouts/sit-ins, minor physical, breaking other school rules, profanity/obscenity, tobacco, ignoring school personnel, no id, unauthorized parking, non-educational computer use
	3	Disrupting bus, gambling, fighting (no injury), forgery, plagiarism, display of gang affiliation, bullying, obscenity/profanity/harassment with bias, second occurrence of level 1 or 2, disruptive use of cell phone, JROTC uniform violations, seriously disruptive internet use, other seriously disruptive behavior
Severe	4	Extortion, assault, vandalism, battery, theft, false fire alarm (no evac), fireworks, trespassing, spreading computer viruses, weapon possession, alcohol, physical contact with staff during fight, other very seriously disruptive behavior
	5	Aggravated assault, burglary, theft over $150, credible threats & severe bullying, gang activity, inappropriate sexual contact (including consensual), other disruptive illegal behavior, sexual harassment, false fire alarm (with evac), repeated weapon possession, battery with injury, cyber bullying, hacking school networks, vandalism over $500, drugs, repeated alcohol, mob action
	6	Firearm or lookalike, use of any weapon to harm, crashing school network, arson, bomb threat, robbery, drug or alcohol sale, repeated drug possession, theft over $1000, sexual assault, aggravated battery, kidnapping, attempted murder, murder

See http://cps.edu/Pages/StudentCodeofConduct.aspx for additional documentation.

Table 2: Analytic Sample Characteristics, by Student Suspension Risk

	All Students	Students at Low Risk of Suspension	Medium Suspension Risk	Students at High Risk of Suspension
Sample size	508,454	290,798	144,378	72,955
Black	46%	30%	62%	78%
Latinx	43%	54%	32%	18%
Other	11%	16%	6%	3%
Male	51%	46%	54%	61%
Special education	15%	12%	18%	26%
Free/reduced lunch	85%	83%	87%	90%
History of abuse or neglect	8%	4%	10%	16%
Prior residential moves	2.1	1.8	2.3	2.7
Prior school moves	2.4	2.2	2.7	3.1
One or more severe infraction(s)	9%	3%	11%	25%
Received one or more ISS	14%	8%	19%	32%
Received one or more OSS	21%	9%	30%	55%
Total days ISS	0.46	0.17	0.61	1.34
Total days OSS	1.40	0.40	1.81	4.56

Each observation is for a single student in a given year. Suspension risk is modeled using 2007 data to predict 2008 suspensions, and then applying that model to future years. Low-risk students have a predicted propensity below 0.2; high-risk have a predicted propensity above 0.5. Prior school moves include the transition to high school, accounting for the fact that some students are in schools that serve both middle and high school grades and do not transition schools after grade 8. Student addresses are coded at the student's Census block group. Therefore, residential moves within the same block group are not observed in the data. In addition, we only observe student addresses twice a year. As such, the number of prior residential moves is likely an undercount of residential moves.

Table 3: Regression of Treatment Measure on Recent Events

	Percent of Severe Events Receiving Suspension, period t *(t = 2 weeks)*
Lagged Severe Events, t-1	-0.152
	(0.42)
Lagged Minor Events, t-1	0.007
	(0.06)
Lagged Severe Events, t-2	0.39
	(0.42)
Lagged Minor Events, t-2	-0.0233
	(0.06)
Lagged Severe Events, t-3	-0.516
	(0.41)
Lagged Minor Events, t-3	-0.015
	(0.06)
Lagged Severe Events, t-4	0.371
	(0.41)
Lagged Minor Events, t-4	-0.040
	(0.06)
N	5,921

+ $p<0.010$ * $p<0.05$ ** $p<0.01$
Regression includes school and time period fixed effects. Observations are at the school-time period level, and missing in periods without severe events. Events are measured per enrolled student, with coefficients scaled by 1000 for readability.

Table 4: Overall Results

	(1)	(2)	(3)	(4)
Standardized Test Scores				
Pr($Suspension|Severe$)	-0.12*	-0.14*	-0.22**	-0.13*
	(0.056)	(0.054)	(0.066)	(0.054)
Sample Size	218,138	216,771	216,848	215,907
Adjusted Percent Attendance				
Pr($Suspension|Severe$)	-5.6	-5.6	-5.4*	-3.5*
	(3.6)	(3.5)	(2.7)	(1.5)
Sample Size	433,297	427,149	427,390	425,357
Perceived Safety				
Pr($Suspension|Severe$)	-0.10	-0.09	0.007	-.09
	(0.14)	(0.14)	(0.17)	(0.097)
Sample Size	214,549	213,434	213,516	212,102
Model				
School F. E.	X	X	X	X
Time Var. School Controls	X	X	X	X
Student Characteristics		X		
Student Time Var. Controls		X	X	X
Student F. E.			X	X
Treatment at Grade Level				X

+ p<0.10 * p<0.05 ** p<0.01

Each estimate arises from a separate regression. Standard errors clustered at school level. Observations are at the student-year level. Specification (4) is the same as (3), with treatment defined at the grade rather than school level. All models control for disciplinary infractions. Test score and attendance models control for student and teacher reports of safety and classroom misbehavior. Test scores are standardized at the grade level and are averages of math and reading scores. Adjusted attendance is measured percent of enrolled days (less suspended days) that the child was present in school. Negative coefficients imply improved test scores when suspension for severe infractions is reduced.

Table 5: Results by School Characteristics

	Test Scores	Attendance	Safety
Pr($Suspension|Severe$) X			
High Baseline Suspension Use	-0.13	-15.0*	-0.31
	(0.14)	(7.2)	(0.28)
Moderate Baseline Use	-0.13*	-1.4	0.18
	(0.059)	(3.8)	(0.24)
Low Baseline Suspension Use	-0.39***	-5.5	0.14
	(0.10)	(4.8)	(0.29)
Sample Size	212,689	421,703	209,174
p-value Low vs. High	0.039	0.26	0.27
Pr($Suspension|Severe$) X			
> 75% African American	0.062	-15.0*	-0.50
	(0.13)	(7.0)	(0.37)
>75% Latinx	-0.19**	1.0	0.46*
	(0.07)	(4.0)	(0.21)
Racially Diverse	-0.34***	-6.0+	-0.26
	(0.09)	(4.0)	(0.22)
Sample Size	216,848	427,390	213,516
p-value AA vs. Lt	0.09	0.05	0.03
p-value AA vs. RD	0.01	0.27	0.55
p-value Lt vs. RD	0.18	0.16	0.02

+ p<0.10 * p<0.05 ** p<0.01 *** p<0.001

Standard errors clustered at school level. Observations are at the student-year level. All columns show results from model (3), with student fixed effects, treatment at the school level, and time-varying controls for school population, student characteristics, disciplinary infractions, as well as student and teacher reports of safety and classroom misbehavior. Negative coefficients imply improved perceptions of safety when suspension for severe infractions is reduced.

Table 6: Results by Student Risk

	Test Scores	Attendance	Safety
Pr($Suspension|Severe$)	-0.21**	-5.7*	0.043
	(0.066)	(2.8)	(0.18)
Pr($Suspension|Severe$) X OSS Risk	-0.01	-0.04	-0.15*
	(0.04)	(0.8)	(0.06)
Sample Size	132,243	427,149	213,434
Pr($Suspension|Severe$)	-0.22**	-5.8*	0.02
	(0.066)	(2.8)	(0.18)
Pr($Suspension|Severe$) X Severe Infraction Risk	0.10*	0.47	-0.12+
	(0.05)	(0.01)	(0.064)
Sample Size	132,211	426,972	213,363

+ p<0.10 * p<0.05 ** p<0.01 *** p<0.001

Standard errors clustered at school level. Observations are at the student-year level. All columns show results from model (3), with student fixed effects, treatment at the school level, and time-varying controls for school population, student characteristics, disciplinary infractions, as well as student and teacher reports of safety and classroom misbehavior. Risk variables range from 0 to 1. Negative coefficients imply improved perceptions of safety when suspension for severe infractions is reduced.